

Relatório Final de Estágio

Mestrado Integrado em Medicina Veterinária

**ESTUDO DE ASSOCIAÇÃO GENÓMICA: PRODUÇÃO DE GORDURA
NO LEITE.**

Fernando José Varandas de Almeida

Orientador:

Prof. Dr. Júlio Gil Vale Carvalheira

Coorientador:

Dr. Albano Gonçalo Beja Pereira

Porto 2013

Relatório Final de Estágio

Mestrado Integrado em Medicina Veterinária

**ESTUDO DE ASSOCIAÇÃO GENÓMICA: PRODUÇÃO DE GORDURA
NO LEITE.**

Fernando José Varandas de Almeida

Orientador:

Prof. Dr. Júlio Gil Vale Carvalheira

Coorientador:

Dr. Albano Gonçalo Beja Pereira

Porto 2013

RESUMO

O presente Relatório Final de Estágio insere-se no 6º ano do Mestrado Integrado em Medicina Veterinária do Instituto de Ciências Biomédicas de Abel Salazar. O meu estágio foi realizado na área de Melhoramento Animal.

O estágio decorreu durante 16 semanas no CIBIO – Centro de Investigação em Biodiversidade e Recursos Genéticos.

Inicialmente, não diretamente relacionada ao tema deste relatório, o estágio consistiu na aquisição de competências práticas mediante o acompanhamento da rotina do laboratório. Eu tive a oportunidade de elaborar um cardeno de laboratório (*laboratory notebook*), bem como de acompanhar à extração de ADN (ácido desoxirribonucleico) de diferentes tipos de amostras (tecidos orgânicos e sémen), a realização de PCR (*Polymerase Chain Reaction*) e a testagem dos resultados através de eletroforese em gel de agarose.

Numa segunda fase do meu estágio houve a introdução da utilização do *software* PLINK o qual permite estudar a existência de associações genéticas (neste caso a quantidade de gordura do leite) e polimorfismos ao nível de simples nucleótidos no genótipo dos animais (SNPs). Os *inputs* para o PLINK tiveram de ser gerados através de dados provenientes da genotipagem de 195 touros (entre 30.000 e 60.000 SNPs por animal) cujo material tinha sido enviado para processamento no laboratório Illumina®. A visualização gráfica dos resultados foi feita com a utilização do programa Haploview.

Tendo terminado o meu estágio considero ter atingido plenamente os objetivos pedagógicos de enriquecimento da formação profissional, integração dos conhecimentos adquiridos durante o curso e a utilização de novas ferramentas informáticas na área da genética no domínio do Melhoramento Animal.

AGRADECIMENTOS

Ad maiorem Dei gloriam

Um especial obrigado ao meu orientador, Professor Júlio Carvalheira, por toda a disponibilidade e atenção prestada. A sua sempre boa disposição e os seus conselhos ajudaram-me muito.
Obrigado.

Para todos do CIBIO, em especial ao meu coorientador, Dr. Albano, e a Vânia.
Ao grupo de informática do Bovinfor, Joel, Adriana e Armando. Muito obrigado pela grande ajuda até nos mais básicos problemas.

Ao meu pai e a minha avó, *in memoriam*, que não podem estar comigo hoje mas cujos ensinamentos levo no coração. Se sou o que sou hoje, agradeço-vos.

Às minhas irmãs, Dulce e Alice, pelas palavras de conforto durante o curso.

À minha mãe e aos meus familiares, sem a vossa ajuda seria impossível.

Aos meus amigos e colegas de curso.

ÍNDICE GERAL

Resumo.....	iii
Agradecimento.....	iv
Introdução.....	1
Melhoramento animal: da antiguidade aos tempos modernos.	
1 – Etapa empírica.....	1
2 – Melhoramento genético animal no século XXI.....	1
3 – Avaliação genómica e seleção assistida por marcadores.....	2
4 – Estudos de Associação Genética.....	4
Materiais e Métodos.....	6
5 – Gestão de Dados	
5.1 – Quais ficheiros criar?.....	6
5.2 – Como foram criados os ficheiros FAM, MAP e LGEN.....	9
6 – A aplicação PLINK para análise de dados.....	10
Resultados.....	12
7 – Controlo da Qualidade dos Dados.....	12
8 – Primeiras estatísticas.....	13
9 – Estudo de associação.....	15
10 – Visualização gráfica dos resultados.....	16
Discussão e Conclusão.....	19
Bibliografia.....	21

INTRODUÇÃO

Melhoramento animal: da antiguidade aos tempos modernos.

1 – Etapa empírica.

Não sabemos quando o ser humano reconheceu a existência da hereditariedade pela primeira vez. Sabemos contudo que os primeiros indícios da domesticação animal ocorreram à cerca de dez mil anos no Crescente Fértil fruto de um processo que se convencionou chamar de Revolução Neolítica. Este processo poderá ter iniciado com a simples subtração de alguns indivíduos selvagens visando o controlo do seu ciclo de vida de forma a obter uma fonte de alimento segura e permanente. A cada geração, essa população submetida às condições de vida e reprodução distintas das populações que permaneciam selvagens, um conjunto de características fenotípicas eram eliminadas em detrimento de outras, originando uma maior homogenização das características fenotípicas desejadas pela redução e fixação de genótipos específicos (Mazoyer & Roudart, 2009).

Podemos dizer então que este foi o princípio do melhoramento animal, ou seja, quando o Homem caçador/coletor converteu-se em cultivador e sedentário criando os seus próprios animais e vegetais sem ter de depender exclusivamente da natureza. Esta ideia totalmente empírica de melhoramento animal baseada na semelhança de indivíduos aparentados resultou em alterações profundas nas características morfológicas e fisiológicas dos animais selecionados. Entretanto poucas sugestões, se houve alguma, foram propostas para explicar estes fenómenos no período pré-histórico (Klug *et al.*, 2010).

A falta de um modelo teórico que explicasse a herança genética abrandou o melhoramento animal por muitos anos, todavia o melhoramento genético animal estabelecido em princípios científicos começou a tornar-se realidade através de uma série de fatos ocorridos nos finais do século XIX e durante o século XX. A redescoberta dos princípios de Mendel, o desenvolvimento dos princípios da genética quantitativa, a criação de associações e cooperativas, o desenvolvimento da inseminação artificial, a descoberta da estrutura do ADN (ácido desoxirribonucleico) entre outros fatores permitiram uma revolução nos programas de seleção no melhoramento animal.

2 – Melhoramento genético animal no século XXI.

No final do século XIX a Revolução Industrial veio a transformar a produção animal deixando de ser atividade apenas de subsistência e extrativista para ser conduzida como uma atividade comercial (Coutinho & Rosário, 2010). Houve nesse contexto uma procura crescente por animais que apresentassem melhor desempenho e que fossem melhor adaptados às diversas condições ambientais. O melhoramento tradicional, fundamentado na teoria da genética quantitativa, assegurou a maior parte do progresso genético obtido nas características produtivas (Ledur *et al.*, 2003).

Contudo esta forma de seleção, ao depender da recolha de informação produtiva, requer tempo e precisão com que se fazem as estimativas, além de diminuir de eficiência com características de baixa heritabilidade. Mas com a descoberta da estrutura do ADN, o desenvolvimento da tecnologia do ADN recombinante e ao advento dos marcadores moleculares, vislumbrou-se a possibilidade de incrementar a informação oriunda do fenótipo com aquela proveniente diretamente do ADN, permitindo que os processos de identificação e seleção dos animais com genótipos superiores pudessem ser realizados de forma mais rápida e eficiente (Coutinho & Rosário, 2010).

Neste sentido o recente sequenciamento dos genomas de animais domésticos, o desenvolvimento de programas de bioinformática e o uso integrado das diferentes estratégias de melhoramento deverão permitir a identificação de novos genes de interesse económico e, desta forma, permitir avanços significativos no uso de informações moleculares para o melhoramento animal (Coutinho *et al.*, 2007).

3 – Avaliação genómica e seleção assistida por marcadores.

As características produtivas, isto é, aquelas que apresentam maior interesse económico, são normalmente características quantitativas e são influenciadas por vários genes em diferentes *loci*. Assume-se que o efeito individual de cada gene associado a essa característica seja pequeno (Regitano *et al.*, 2012), e este raciocínio tem sido a base dos modelos matemáticos de genética quantitativa aplicados no melhoramento animal. Entretanto, sabe-se que alguns desses genes podem ter um efeito maior que o esperado na expressão fenotípica que outros (Montaldo & Meza-Herrera, 1998), e por isso tais genes são chamados de genes maiores (*major*) de uma característica de *loci* quantitativo (QTL, do inglês *Quantitative Trait Loci*).

Assim, embora o termo QTL possa ser aplicado à todos os genes envolvidos na expressão de características quantitativa, na prática refere-se somente aos genes maiores (*major*) dado que só a estes interessa detectar e mapear. A ideia então por detrás da seleção assistida por

marcadores (do inglês *Marker-Assisted Selection*) é que os genes (ou regiões) com um efeito mais significativo na expressão fenotípica possam ser identificados e serem alvo de seleção mais intensa (Eenennaam, 2004). De facto, os marcadores para características complexas (quantitativas) estão associados a somente um dos muitos genes que contribuem para a expressão fenotípica, e ao seleccionar animais portadores do marcador com um impacto positivo permite o melhoramento mais rápido da característica (Eenennaam, 2004).

Dentro deste contexto a definição de marcador genético é um gene, um segmento de ADN ou mesmo um SNP identificável no genoma e que pode ser utilizado para seleção pela sua proximidade a um QTL (Van der Werf, 2000).

Facilmente se compreende, do que anteriormente foi dito, que a incorporação do conhecimento do genótipo do indivíduo para a sua avaliação parece ser o futuro dos programas de seleção. De facto, todos os jovens touros comprados pelas principais organizações de inseminação artificial nos Estados Unidos da América são seleccionados com base nesta avaliação (Wiggans *et al.*, 2010).

A avaliação genómica, ao permitir identificar indivíduos portadores de marcadores associados à QTL de interesse, surge como complemento as avaliações clássicas (genealógicas e produtivas). O conjunto destas duas novas ferramentas da genética, isto é, a avaliação genómica e a seleção assistida por marcadores, apresentam-se como métodos que permitem avaliar de forma mais eficaz os animais a nível genético desde a sua juventude, permitindo aumentar a eficiência e acurácia da seleção, diretamente relacionados ao ganho genético (Davis & Denise, 1998).

Não obstante o objetivo deste Relatório Final de Estágio não ser a avaliação genómica, fez-se contudo um estudo de associação genética entre um fenótipo (gordura no leite calculado por métodos clássicos de genética quantitativa) e um marcador genético (SNP) o qual está inserido num projeto de Seleção Genómica em Populações de Bovinos Leiteiros em curso no CIBIO.

Para tal 74 touros de uma população Holstein foram genotipados. O ADN foi extraído nos laboratórios do CIBIO (Centro de Investigação em Biodiversidade e Recursos Genéticos) a partir de palhinhas com sémen. As amostras foram posteriormente enviada ao laboratório Illumina® (*Illumina Inc.*, Califórnia, Estados Unidos da América). A genotipagem massiva de SNPs consiste na utilização de um *microchip* - *BovineSNP50 v2 BeadChip* (Figura 1), que

permite a genotipagem de 54.609 SNPs espalhados mais ou menos uniformemente pelo genoma bovino.

A escolha deste *chip* residiu no facto de o mesmo estar comprovadamente capacitado para avaliação do mérito genético, estudos de QTL, de genética comparativa, bem como outros estudos (Anónimo, 2012).

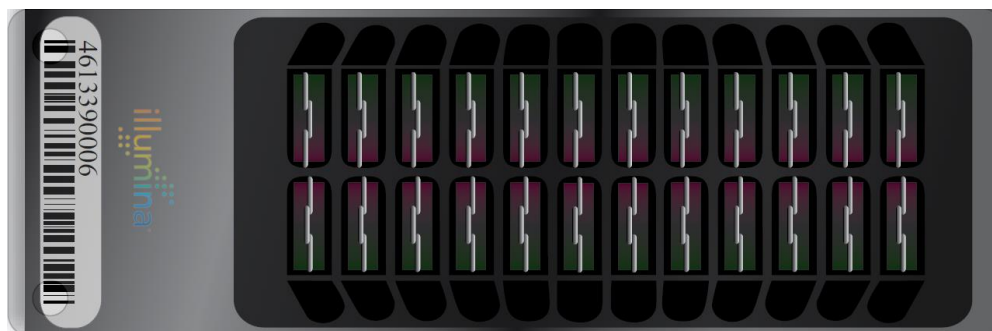


Figura 1. Ilustração de um BovineSNP50 v2 BeadChip.

Deste mesmo laboratório foi-nos fornecido a genotipagem de outros 121 animais, os quais para efeitos comparativos foram adicionados à população Holstein inicial, totalizando 195 animais genotipados. Estes 121 animais não nos foram aleatoriamente fornecidos, com efeito, estes foram por nós solicitados por estarem relacionados com a população Holstein Portuguesa e por terem sido utilizados massivamente na inseminação artificial desta população.

4 – Estudos de Associação Genética.

Os estudos de associação genética são utilizados para identificar genes candidatos ou regiões no genoma que contribuem para uma característica fenotípica. Nestes estudos avaliam-se as associações entre cada polimorfismo simples (SNP) e o fenótipo observado em toda a população. Assim, a maior frequência de um SNP num conjunto de animais cujo fenótipo se destaca da população pode ser interpretado como relação entre os dois, embora outras interpretações possa ser igualmente válidas (Lewis & Knight, 2012). De facto, a interpretação da associação genética pode ter três formas: (1) associação direta, na qual o SNP genotipado é a variante causal da característica; (2) associação indireta, na qual o SNP está em desequilíbrio de *linkage* com a verdadeira variante responsável pela característica; (3) um falso positivo no qual o resultado aparece por erro aleatório ou sistemático (Lewis & Knight, 2012).

Para este estudo foi utilizado o programa informático PLINK. Este programa foi desenvolvido por Shaun Purcell (2007), no CHGR (*Center for Human Genetic Research*), do

Hospital Geral de Massachusetts (*Massachusetts General Hospital* - MGH), com apoio da equipa de investigadores do próprio MGH, e pode ser obtido através da *internet* pelo endereço: <http://pngu.mgh.harvard.edu/purcell/plink/>.

A origem do PLINK está relacionada com a necessidade de possuir ferramentas de análise genética capazes de lidar com um grande número de dados genómicos de forma computacionalmente eficiente (Purcell *et al.*, 2007). Assim, é possível analisar uma grande quantidade de dados provenientes da genotipagem de marcadores para muitos indivíduos. Outra vantagem na utilização da aplicação PLINK resulta do facto de esta poder ser utilizada para dados genómicos de qualquer espécie de organismo diplóide.

Além dos argumentos anteriores, a sua escolha como ferramenta para este Relatório Final de Estágio foi o facto de possuir uma estrutura clara e intuitiva de utilização dos seus comandos, facilmente aprendido por qualquer utilizador pouco experiente.

MATERIAL E MÉTODOS

5 – Gestão de Dados.

5.1 – Quais ficheiros criar?

Os dados provenientes do laboratório Illumina® vieram em diferentes ficheiros de texto com extensão “.txt”, estes ficheiros não estavam prontos para a utilização no PLINK sendo, por isso, modificados para criar ficheiros de entrada (*inputs*) reconhecíveis pelo programa.

Neste contexto foram criados 3 ficheiros de textos com base nos resultados laboratoriais: um ficheiro FAM, um ficheiro MAP e um ficheiro LGEN. Estes três ficheiros anteriores são chamados *long-format files*, podem ser utilizados pela aplicação PLINK e serão descritos abaixo.

O ficheiro FAM possui 6 colunas separadas por tabulação, identificadas a seguir na Tabela 1.

Coluna	Descrição
1	Identificação da Família ou Grupo dos Indivíduos
2	Identificação do Indivíduo
3	Identificação do Pai
4	Identificação da Mãe
5	Sexo do Indivíduo; (1=macho; 2=fêmea)
6	Fenótipo do Indivíduo; (-9=sem informação; 1=não afetado; 2=afetado)

Tabela 1.

Este ficheiro corresponde a identificação de todos os indivíduos presente no estudo. É apresentado como uma lista onde cada linha corresponde a um indivíduo. Um exemplo do conteúdo do ficheiro FAM é visto na Figura 2.

27	128	27	323	1	-9
27	257	27	440	1	-9
27	268	27	440	1	-9
27	360	27	0	2	-9
28	126	28	321	1	-9
29	127	29	322	1	-9

Figura 2.

A combinação da identificação da família (coluna 1) e da identificação do indivíduo (coluna 2) é única, e podem ser apenas números ou uma combinação alfanumérica. A última coluna – coluna 6 – correspondente ao fenótipo e pode ser um valor qualitativo (1 ou 2) ou quantitativo (exemplo: 3.6908), separando a unidade das casas decimais por ponto.

O ficheiro MAP, por sua vez, é também um ficheiro de texto separado por tabulação, contudo este possui informação sobre os SNPs. Neste caso temos 4 colunas onde cada uma está identificada na tabela a seguir.

Coluna	Descrição
1	Cromossoma onde se situa o SNP (1 a 29, X, Y e 0 se não localizado)
2	Identificação (nome) do SNP
3	Distância Genética (morgans)
4	Posição do Par de Nucleótidos

Tabela 2.

O ficheiro MAP mapeia os diferentes SNPs pelos cromossomas. A coluna 1 do ficheiro MAP diz qual o cromossoma que o SNP pertence, podendo ser um valor entre 1 e 29 (correspondente aos 29 pares de cromossomas autossómicos dos bovinos), um valor X ou Y, caso seja um cromossoma sexual, ou o valor 0 (zero), se não foi possível determinar o cromossoma.

A coluna 2 identifica o nome que se atribui ao SNP.

A coluna 3 corresponde à distância genética, e a sua definição corresponde a medida da diferença de material genético entre diferentes espécies ou indivíduos da mesma espécie (Kalinowski, 2002). De acordo com Purcell (2010) esta medida é mais importante para o conjunto de análises que procuram segmentos partilhados entre indivíduos. Para os testes de associação básicos os valores de distância genética podem ser zero.

A coluna 4 corresponde a posição, em pares de base, da localização do SNP no cromossoma, e deverão ser sempre valores inteiros positivos.

Um exemplo do conteúdo do ficheiro MAP é:

1	BovineHD0100010456	0	36524125
1	BovineHD0100010461	0	36547637
1	Hapmap50634-BTA-26738	0	36595968
1	ARS-BFGL-NGS-44057	0	36636902
1	ARS-BFGL-NGS-38419	0	36664556

Figura 3. Exemplo do conteúdo do ficheiro MAP.

Por último o ficheiro LGEN é um ficheiro de texto que fornece a informação dos alelos de cada SNP para cada indivíduo genotipado. Trata-se portanto de um ficheiro muito comprido onde para cada indivíduo há dezenas de milhares de linhas correspondentes aos seus alelos para cada SNP.

O ficheiro LGEN possui 5 colunas. A descrição de cada coluna é feita na Tabela 3, e tal como os anteriores cada coluna também é separada por tabulação.

Coluna	Descrição
1	Identificação da Família (semelhante ao ficheiro FAM)
2	Identificação individual (semelhante ao ficheiro FAM)
3	Identificação (Nome) do SNP
4	Alelo 1 deste genótipo
5	Alelo 2 deste genótipo

Tabela 3.

As colunas 1 e 2 são a identificação do indivíduo, e a coluna 3 corresponde o nome do SNP para qual o indivíduo foi genotipado. As colunas 4 e 5 são representadas pelas letras A, G, C e T representando aos nucleótidos adenina, guanina, citosina e timina de cada alelo, respectivamente.

Desta forma temos na Figura 4 um exemplo do ficheiro LGEN.

56	179	ARS-BFGL-BAC-10972	G	G
56	179	ARS-BFGL-BAC-10986	C	C
56	179	ARS-BFGL-BAC-10993	A	A
56	179	ARS-BFGL-BAC-11000	C	C
56	179	ARS-BFGL-BAC-11003	A	A
56	179	ARS-BFGL-BAC-11039	G	G

Figura 4. Exemplo do conteúdo do ficheiro LGEN.

5.2 – Como foram criados os ficheiros FAM, MAP e LGEN.

Como dito anteriormente os dados do laboratório americano não vieram prontos a serem utilizados no programa PLINK. Desta forma estes foram editados para produzir os ficheiros requeridos.

Os resultados laboratoriais vieram separados em três ficheiros diferentes: um *FinalReport*, um *Sample_map* e um *SNP_map*.

O *SNP_map*, cujo exemplo do seu conteúdo pode ser visto na Figura 5, permite a criação do ficheiro MAP. Para a sua produção utilizou-se um simples programa para escritório (*Office suite*) denominado *LibreOffice*[®], no qual, a partir do *SNP_map*, eliminou-se as colunas não necessárias deixando apenas aquelas que permitiam criar o ficheiro MAP, isto é, as colunas “*chromosome*”, “*name*” e “*position*”, respectivamente.

Na Figura 5, a seguir, se encontra representado um exemplo do conteúdo de um ficheiro *SNP_map*.

Index	Name	Chromosome	Position	GenTrain	Score	SNP	ILMN	Strand
9	ARS-BFGL-BAC-10919	14	31267746	0.8255	[A/G]	TOP	TOP	2
10	ARS-BFGL-BAC-10951	10	17911906	0.9056	[T/G]	BOT	BOT	2
11	ARS-BFGL-BAC-10952	10	18882288	0.9184	[A/G]	TOP	TOP	2
12	ARS-BFGL-BAC-10960	10	20609250	0.9205	[A/G]	TOP	TOP	2
13	ARS-BFGL-BAC-10972	10	20792754	0.8432	[G/C]	BOT	BOT	102
14	ARS-BFGL-BAC-10975	10	21225382	0.7991	[A/G]	TOP	TOP	2
15	ARS-BFGL-BAC-10986	10	26527257	0.8941	[A/C]	TOP	BOT	2
16	ARS-BFGL-BAC-10993	10	78512500	0.8649	[A/G]	TOP	BOT	2
17	ARS-BFGL-BAC-11000	10	79252023	0.9433	[T/G]	BOT	BOT	2
18	ARS-BFGL-BAC-11003	10	80410977	0.8842	[T/C]	BOT	BOT	2

Figura 5.

O ficheiro *FinalReport* forneceu os dados para a produção do ficheiro LGEN. Um exemplo do seu conteúdo pode ser visto na Figura 6.

SNP Name	Sample ID	Allele1 - Forward	Allele2 - Forward	Allele1 - Top	Allele2 - Top	GC
ARS-BFGL-BAC-10172	34459773	G	G	G	G	0.9081
ARS-BFGL-BAC-1020	34459773	A	G	A	G	0.8939
ARS-BFGL-BAC-10245	34459773	C	C	G	G	0.7996
ARS-BFGL-BAC-10345	34459773	A	C	A	C	0.9458
ARS-BFGL-BAC-10365	34459773	-	-	-	-	0.0000
ARS-BFGL-BAC-10375	34459773	G	G	G	G	0.8746
ARS-BFGL-BAC-10591	34459773	A	A	A	A	0.8587
ARS-BFGL-BAC-10793	34459773	C	C	G	G	0.8640
ARS-BFGL-BAC-10867	34459773	C	C	G	G	0.9326
ARS-BFGL-BAC-10919	34459773	A	A	A	A	0.8746
ARS-BFGL-BAC-10951	34459773	-	-	-	-	0.0000
ARS-BFGL-BAC-10952	34459773	A	G	A	G	0.9498

Figura 6.

Tal como o ficheiro anterior, eliminou-se as colunas não desejadas deixando apenas aquelas colunas que permitiam criar o ficheiro LGEN descrito na Tabela 3.

Neste estudo foi utilizado como referência para os alelos de cada SNP os *Allele1 - Top* e *Allele2 - Top*. A escolha destes alelos teve como base a indicação no manual de utilização do programa GCTA (Yang *et al.*, 2010) que, semelhante a aplicação PLINK, é um programa que permite estudos de associação do genoma completo (*Genome-wide Association Study*).

O último ficheiro a ser criado foi o ficheiro FAM. Este foi feito com base nas informações contida no *Sample_map* (identificação dos animais genotipados) e posteriormente formação de um *pedigree* (com a identificação dos progenitores). Para identificar os progenitores dos animais genotipados foi feita a busca no Bovinfor, que corresponde a uma base de dados nacional sobre a informação produtiva, reprodutiva e genealógica do gado bovino leiteiro. O critério para formação do *pedigree* foram as linhas paternas, sendo o pai (macho) mais antigo como referência para determinar a família. A construção de *pedigree* foi feita com recurso ao MATLAB® (2010) e SAS® (1990).

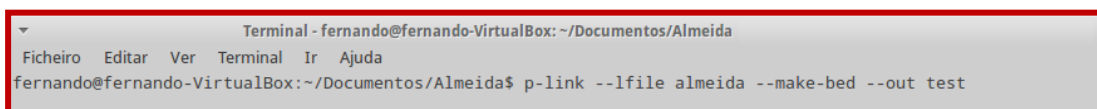
6 – A aplicação PLINK para análise de dados.

A aplicação PLINK trata-se de um *software* desenvolvido sob a Licença Pública Geral (*General Public License - GPL*) sendo portanto de código fonte e uso livre, e pode ser executado em diferentes sistemas operacionais. No meu caso foi instalado num portátil com Linux Ubuntu, versão 12.04.2 LTS e arquitetura 32 bits. A versão utilizada da aplicação PLINK para a realização deste Relatório de Estágio foi v1.07 de 10 de Agosto de 2009.

Este programa informático foi escrito em C/C++ e a sua utilização requer a introdução de ordens pela linha de comandos do terminal. Dito de outra forma, o funcionamento do PLINK faz-se ao digitar comandos (iniciados pelos caracteres “- -”) no terminal do Linux para especificar os ficheiros de dados e/ou métodos a serem utilizados. Os resultados são salvos em ficheiros cuja extensão está relacionada ao comando utilizado.

A utilização de interface com linhas de comando (CLI – do acrónimo inglês *Command-Line Interface*) implica que o utilizador saiba para que serve cada comando e os seus respectivos parâmetros. Não é do objetivo deste relatório introduzir as noções básicas, contudo estas são facilmente encontrada na *internet*. Entretanto é relevante desenvolver uma breve introdução ao uso da aplicação PLINK.

Um exemplo da linha de comando para utilização da aplicação PLINK seria:



```

Terminal - fernando@fernando-VirtualBox: ~/Documentos/Almeida
Ficheiro Editar Ver Terminal Ir Ajuda
fernando@fernando-VirtualBox:~/Documentos/Almeida$ p-link --lfile almeida --make-bed --out test
  
```

Figura 7.

Onde *p-link* corresponde a indicação para utilização do programa PLINK, *--lfile* corresponde a instrução para leitura de três ficheiros em formato longo (do inglês *long-format files*), a opção *--make-bed* produz um ficheiro binário e mais outros dois ficheiros tendo, por fim, o comando *--out test* que indica que os ficheiros produzidos terão como nome *test* seguido da extensão resultante dos comandos. Neste caso então teremos quatro novos ficheiros: *test.bim*, *test.fam*, *test.bim* mais o *test.log*.

A Tabela 4 mostra alguns exemplos dos tipos de ficheiros de saída (*output*):

Extensão do ficheiro	Comando utilizado	Descrição
bed	--make-bed	Ficheiro binário
freq	--freq	Tabela de frequência alélica
assoc	--assoc	Resultados de associação
log	(não é necessário)	Lista comandos utilizados
lmiss	--missing	Estatísticas de não-resposta

Figura 4.

RESULTADOS

7 – Controlo da Qualidade dos Dados.

Antes de iniciar as primeiras análises efectuou-se o controlo da qualidade dos dados. A Figura 8 corresponde ao ficheiro *log* que obteve.

```
@-----@
|          PLINK!          |          v1.07          |          10/Aug/2009          |
|-----|-----|-----|
| (C) 2009 Shaun Purcell, GNU General Public License, v2 |
|-----|-----|-----|
| For documentation, citation & bug-report instructions: |
| http://pngu.mgh.harvard.edu/purcell/plink/              |
|-----|-----|-----|
@-----@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

Writing this text to log file [ almeida_binario.log ]
Analysis started: Tue Apr 30 10:44:50 2013

Options in effect:
  --lfile almeida
  --maf 0.02
  --geno 0.1
  --mind 0.2
  --hwe 0.000001
  --nonfounders
  --cow
  --make-bed
  --out almeida_binario
```

Figura 8.

corresponde ao limiar máximo de desequilíbrio de Hardy-Weinberg para cada SNP. Enquanto que os outros comandos: *cow*, *nonfounders* e *make-bed*, identificam o gênero bovino, a inclusão de todos os indivíduos na análise e a produção de um ficheiro binário, respectivamente.

Com estas opções os resultados obtidos foram os vistos na Figura 9.

```
54609 (of 54609) markers to be included from [ almeida.map ]
Reading pedigree information from [ almeida.fam ]
453 individuals read from [ almeida.fam ]
0 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
0 cases, 0 controls and 453 missing
281 males, 172 females, and 0 of unspecified sex
Before frequency and genotyping pruning, there are 54609 SNPs
257 founders and 196 non-founders found
Writing list of removed individuals to [ almeida_binario.irem ]
303 of 453 individuals removed for low genotyping ( MIND > 0.2 )
4821 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ almeida_binario.hh ]
40 markers to be excluded based on HWE test ( p <= 1e-06 )
  0 markers failed HWE test in cases
  0 markers failed HWE test in controls
Total genotyping rate in remaining individuals is 0.899763
9937 SNPs failed missingness test ( GENO > 0.1 )
9781 SNPs failed frequency test ( MAF < 0.02 )
After frequency and genotyping pruning, there are 38450 SNPs
After filtering, 0 cases, 0 controls and 150 missing
After filtering, 150 males, 0 females, and 0 of unspecified sex
```

Figura 9

Como pode ser visto na figura ao lado os comandos utilizados foram o *mind*, que corresponde a proporção máxima da não-resposta (*missingness*) da genotipagem por indivíduo, o *MAF* (*Minor Allele Frequency*) que corresponde ao valor da frequência do alelo menos frequente de cada SNP, e por último o comando *geno* que depura os SNPs que apresentam uma proporção de não-resposta (*missingness*) superior ao valor escolhido – *0.1*. O comando *hwe*

Da leitura da figura ao lado temos: 453 animais, sendo 281 machos e 172 fêmeas. Contudo, depois de removidos aqueles indivíduos com baixa genotipagem (*low genotyping*) restaram apenas 150 touros. Dos 54.609 SNPs iniciais contidos no ficheiro MAP (*almeida.map*) ficaram 38.450 SNPs.

8 – Primeiras estatísticas.

Tendo aplicado os critérios de inclusão aos ficheiros iniciais obtiveram-se as primeiras estatísticas. Os próximos resultados não têm todos necessariamente a ver com o estudo de associação, contudo serão apresentados já que a sua manipulação e interpretação foram um dos objetivos pedagógico do meu estágio.

Assim com a opção *--missing*, que fornece as estatísticas de não-resposta (*missing data*), temos o ficheiro *log* (Figura 10):

```
Writing this text to log file [ miss_stat.log ]
Analysis started: Tue Apr 30 10:58:42 2013

Options in effect:
  --bfile almeida_binario
  --missing
  --out miss_stat

Reading map (extended format) from [ almeida_binario.bim ]
38450 markers to be included from [ almeida_binario.bim ]
Reading pedigree information from [ almeida_binario.fam ]
150 individuals read from [ almeida_binario.fam ]
0 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
0 cases, 0 controls and 150 missing
150 males, 0 females, and 0 of unspecified sex
Reading genotype bitfile from [ almeida_binario.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 38450 SNPs
0 founders and 150 non-founders found
162456 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ miss_stat.hh ]
38450 SNPs with no founder genotypes observed
```

Figura 10.

Desde mesmo comando geram-se outros dois ficheiros que fornecem os resultados de não-resposta por indivíduo (*miss_stat.imiss*) e por SNP (*miss_stat.lmiss*). Um exemplo do conteúdo dos dois ficheiros está contido na Figura 11.

FID	IID	MISS_PHENO	N_MISS	N_GENO	F_MISS	CHR	SNP	N_MISS	N_GENO	F_MISS
18	261	Y	1533	38450	0.03987	0	ARS-BFGL-BAC-13205	8	150	0.05333
18	262	Y	1736	38450	0.04515	0	ARS-BFGL-BAC-33662	0	150	0
17	263	Y	2069	38450	0.05381	0	ARS-BFGL-NGS-100012	8	150	0.05333
15	265	Y	2134	38450	0.0555	0	ARS-BFGL-NGS-100232	2	150	0.01333
20	266	Y	1461	38450	0.038	0	ARS-BFGL-NGS-100341	0	150	0
23	267	Y	1176	38450	0.03059	0	ARS-BFGL-NGS-100372	0	150	0
22	268	Y	1327	38450	0.03451	0	ARS-BFGL-NGS-10051	0	150	0
24	270	Y	2583	38450	0.06718	0	ARS-BFGL-NGS-100549	0	150	0
31	275	Y	1593	38450	0.04143	0	ARS-BFGL-NGS-100941	12	150	0.08

miss_stat.imiss

miss_stat.lmiss

Figura 11.

A correspondência de cada coluna da Figura 11 é: FID – Identificação familiar; IID – Identificação do indivíduo; MISS_PHENO – Ausência de fenótipo (Y=sim, N=não); N_MISS (*imiss*) – Número de SNPs que o indivíduo falhou relativamente aos 38.450 SNPs; N_GENO (*imiss*) – número total de SNPs; N_MISS (*lmiss*) – indivíduos que não possuem o SNP em questão; N_GENO (*lmiss*) – número dos indivíduos analisados; F_MISS – Frequência de

ausência (*missingness*), corresponde a proporção entre (N_MISS e N_GENO); SNP – nome do SNP; CHR – Número do cromossoma.

A seguir foram utilizados comandos para o estudo das frequências alélicas, como pode ser visto na Figura 12.

```
Writing this text to log file [ freq_stat.log ]
Analysis started: Tue Apr 30 11:04:48 2013

Options in effect:
--bfile almeida_binario
--freq
--nonfounders
--out freq_stat

Reading map (extended format) from [ almeida_binario.bim ]
38450 markers to be included from [ almeida_binario.bim ]
Reading pedigree information from [ almeida_binario.fam ]
150 individuals read from [ almeida_binario.fam ]
0 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
0 cases, 0 controls and 150 missing
150 males, 0 females, and 0 of unspecified sex
Reading genotype bitfile from [ almeida_binario.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 38450 SNPs
0 founders and 150 non-founders found
162389 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ freq_stat.hh ]
Writing allele frequencies (all individuals) to [ freq_stat.frq ]
```

Figura 12.

Desde comando foi obtido o ficheiro *freq_stat.frq*, cujo conteúdo pode ser visto parcialmente na Figura 13.

CHR	SNP	A1	A2	MAF	NCHROBS
0	ARS-BFGL-BAC-13205	G	A	0.3838	284
0	ARS-BFGL-BAC-33662	A	G	0.3967	300
0	ARS-BFGL-NGS-100012	A	G	0.1796	284
0	ARS-BFGL-NGS-100232	A	G	0.2061	296
0	ARS-BFGL-NGS-100341	G	A	0.42	300
0	ARS-BFGL-NGS-100372	A	G	0.4933	300
0	ARS-BFGL-NGS-10051	A	G	0.07333	300
0	ARS-BFGL-NGS-100549	G	A	0.4867	300
0	ARS-BFGL-NGS-100941	G	A	0.1304	276
0	ARS-BFGL-NGS-101104	G	A	0.4	300

Figura 13.

Na Figura 13 as colunas A1 representam o nucleótido do alelo *major* (de maior frequência alélica) enquanto a coluna A2 representa o nucleótido para alelo *minor* (de menor frequência alélica). A coluna NCHROBS é o acrónimo para *Number Chromosome Observation*, isto é, o número de cromossomas válidos (*non-missing*). As demais colunas – CHR, SNP e MAF – coincidem com as definições para as figuras anteriores.

9 – Estudo de associação.

O ficheiro FAM, como anteriormente descrito, não possui na sexta coluna um valor para o fenótipo dos animais. Esta coluna está identificada com um valor -9 que o programa PLINK reconhece com ausente (*missing value*). Assim, para o estudo de associação, foram criados dois ficheiros com diferentes valores para o fenótipo. Um primeiro ficheiro, denominado *almeida_quant.phe*, contém valores genéticos da gordura no leite que foram calculados por métodos clássicos de genética quantitativa e foram obtidos na avaliação nacional de Janeiro de 2013. Um segundo ficheiro, nomeado *almeida_1_2.phe*, contém um valor qualitativo para o fenótipo. Neste segundo ficheiro a população total (453 animais) foi dividida em duas subpopulações sendo que os touros portugueses (população Holstein inicial de 74 touros) receberam para o valor de fenótipo o algarismo 2 e os restantes animais tiveram o algarismo 1 como valor para o fenótipo.

Um exemplo do conteúdo dos ficheiros *almeida_quant.phe* e *almeida_1_2.phe* pode ser visto na Figura14.

11	374	1	34	167	19.520394982
84	375	1	35	168	16.012894982
80	376	1	36	169	7.0153949822
79	377	1	37	170	0.4578949822
82	378	1	38	171	15.372394982
89	379	1	39	172	-7.411105018
83	380	1	40	173	7.5338949822
88	381	1	41	174	-16.95760502
2	382	2	42	175	0.6103949822
3	383	2	43	176	-26.13810502
26	384	2	44	177	-3.202105018
5	385	2	45	178	-1.341605018
30	386	2	46	179	4.2703949822
30	387	2	47	180	3.6908949822
34	388	2	48	181	-13.99910502
30	389	2	49	182	31.842394982

almeida_1_2.phe

almeida_quant.phe

Figura 14.

Com estes ficheiros prontos, utilizou-se os seguintes comandos:

Options in effect: --bfile almeida_binario --assoc --adjust --pheno almeida_quant.phe --out as_quantitative	Options in effect: --bfile almeida_binario --assoc --adjust --pheno almeida_1_2.phe --out as_qualitative
--	---

Figura 15.

Tendo-se obtido os resultados presentes nas figuras 16 e 17.

CHR	SNP	BP	NMISS	BETA	SE	R2	T	P
9	Hapmap42751-BTA-98904	57033506	150	9.145	5.118	0.02112	1.787	0.076
9	Hapmap49396-BTA-98905	57069740	150	3.547	2.744	0.01117	1.293	0.1981
9	BTB-01151584	57128694	149	-8.285	3.994	0.02844	-2.074	0.03979
9	BTA-98907-no-rs	57182000	150	-7.114	3.362	0.02936	-2.116	0.03605
9	BTB-01151858	57224923	150	-2.775	2.381	0.009093	-1.165	0.2457
9	BTB-01151930	57288578	150	3.193	2.453	0.01132	1.302	0.1951
9	BTB-01152013	57327067	150	0.9705	2.625	0.0009227	0.3697	0.7121
9	BTB-01152078	57384683	149	-2.27	5.173	0.001309	-0.4389	0.6614
9	BTB-01152124	57412955	150	0.07505	2.784	4.91e-06	0.02696	0.9785
9	ARS-BFGL-NGS-84840	57515862	149	0.3232	2.263	0.0001388	0.1428	0.8866
9	UA-IFASA-4968	57559232	150	-6.189	2.349	0.04479	-2.634	0.009326
9	BTA-83801-no-rs	57616379	150	-5.8	2.837	0.02745	-2.044	0.04273
9	BTB-01427738	57679687	146	2.623	5.043	0.001874	0.52	0.6039

Figura 16. Exemplo do conteúdo do ficheiro obtido. Legenda: CHR – cromossoma; SNP – nome do SNP; BP – localização do SNP em pares de bases; NMISS – número de indivíduos válidos (*non-missing*); BETA – coeficiente de regressão; SE – erro padrão; R2 – coeficiente de determinação; T – valor da distribuição de t; P – valor de p (significância estatística).

A Figura 16 é o resultado para os valores de fenótipo quantitativo, ficheiro *almeida_quant.phe*.

Por sua vez, a Figura 17 são os resultados para os valores de fenótipo qualitativo, ficheiro *almeida_1_2.phe*.

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
3	ARS-BFGL-NGS-37142	124886498	A	0.8871	0.8636	G	0.3619	0.5474	1.241
3	ARS-BFGL-NGS-101315	125045825	G	0.7661	0.7443	A	0.1862	0.6661	1.125
3	ARS-BFGL-NGS-90439	125114116	G	0.7542	0.7216	A	0.3862	0.5343	1.184
3	ARS-BFGL-NGS-35269	125139134	C	0.5306	0.5739	A	0.4775	0.4895	0.8394
3	ARS-BFGL-NGS-30038	125516631	C	0.7258	0.7089	A	0.09818	0.754	1.087
4	Hapmap52120-rs29014886	565142	A	0.6048	0.5263	C	1.71	0.191	1.378
4	Hapmap55620-rs29024432	631320	G	0.75	0.7012	A	0.8375	0.3601	1.278
4	ARS-BFGL-NGS-18108	870437	G	0.7258	0.6705	A	1.048	0.3059	1.301
4	BTB-00281544	993285	G	0.8468	0.9034	A	2.21	0.1372	0.5909
4	ARS-BFGL-NGS-36492	1120424	G	0.9032	0.9253	A	0.4585	0.4983	0.7536

Figura 17. Exemplo do conteúdo. Legenda: CHR – cromossoma; SNP – nome do SNP; BP – localização do SNP em pares de base; A1 - alelo *major*; F_A – frequência do alelo na população nos população casos; F_U – frequência do alelo nos controlos; A2 – alelo *minor*; CHISQ – valor do teste qui-quadrado; P – valor de p; OR – *Odds Ratio*.

10 – Visualização gráfica dos resultados.

Os resultados (*output*) da aplicação PLINK são ficheiros de texto com colunas separadas por espaços. Para a sua visualização recorreu-se ao programa Haploview (Barret *et al.*, 2005), sendo este programa também de código aberto (*open source*) e gratuito.

Com este programa construíram-se gráfico onde se representaram os valor de p, no eixo das ordenadas, relativamente a característica quantitativa (*ficheiro almeida_quant.phe*) e qualitativa (*almeida_1_2.phe*) para cada cromossoma (representado no eixo das abscissas).

Do Figura 18 viu-se que dois cromossomas se destacavam pelo seu valor de p, os cromossomas 3 e 14. Por isso realizaram-se outros quatro gráficos para representar a localização (por pares de base) no eixo das abscissas para os valor de p (eixo da ordenadas) em cada um dos cromossomas individualmente, figuras 19, 20, 22 e 23.

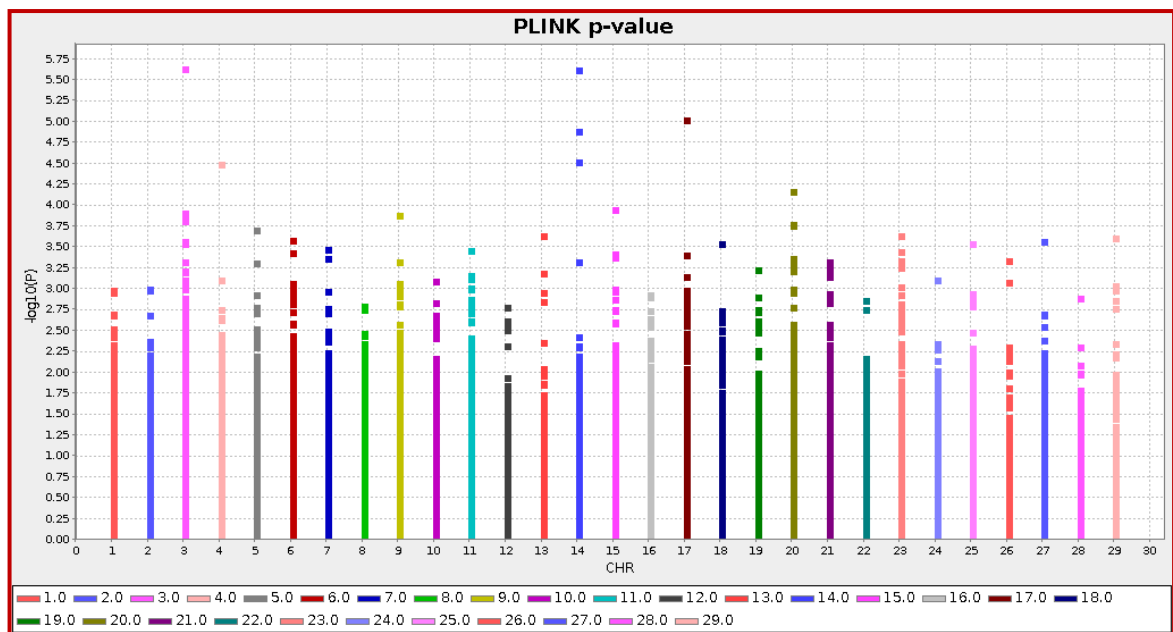


Figura 18. CHR – Cromossoma; $-\log_{10}(P)$ - Corresponde ao negativo do logaritmo de base 10 do valor de p.

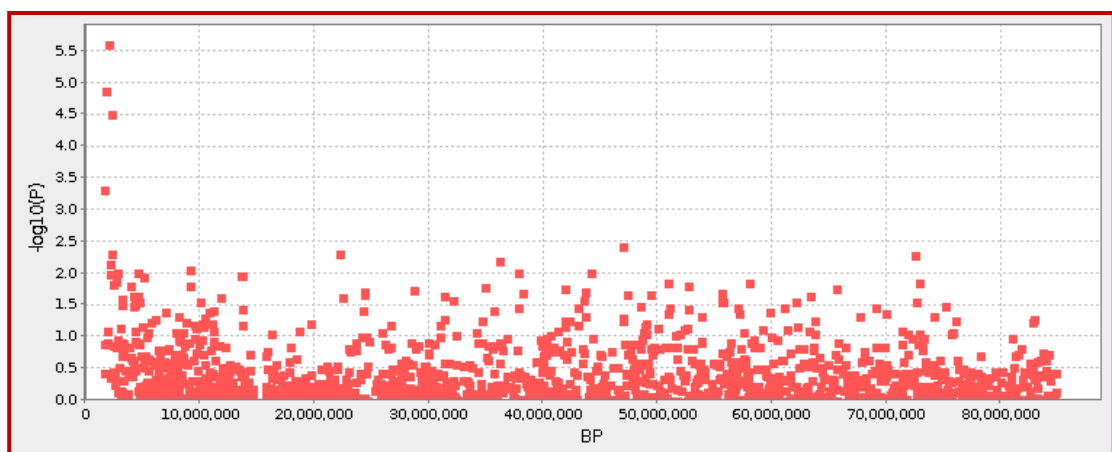


Figura 19. Cromossoma 14. BP – Pares de base; $-\log_{10}(P)$ - Corresponde ao negativo do logaritmo de base 10 do valor de p.

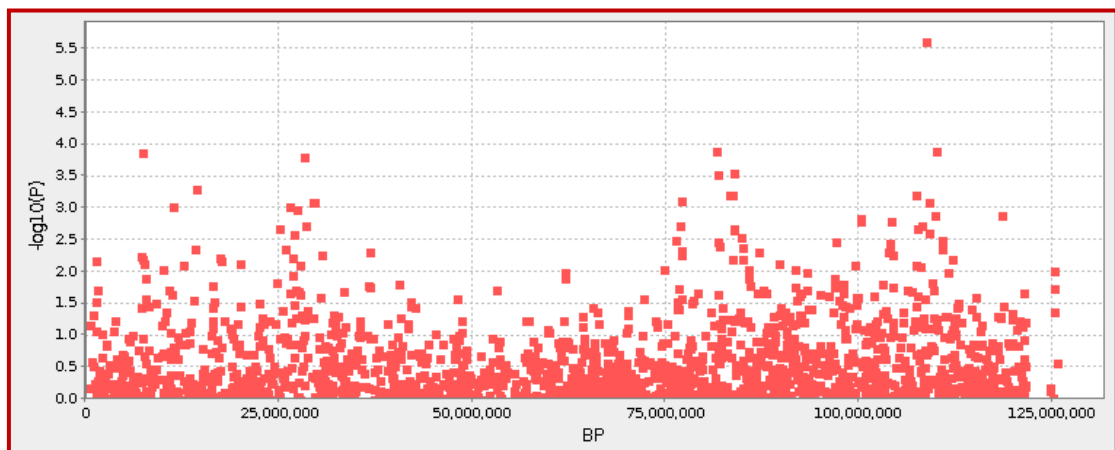


Figura 20. Cromossoma 3. BP – Pares de base; $-\log_{10}(P)$ - Corresponde ao negativo do logaritmo de base 10 do valor de p.

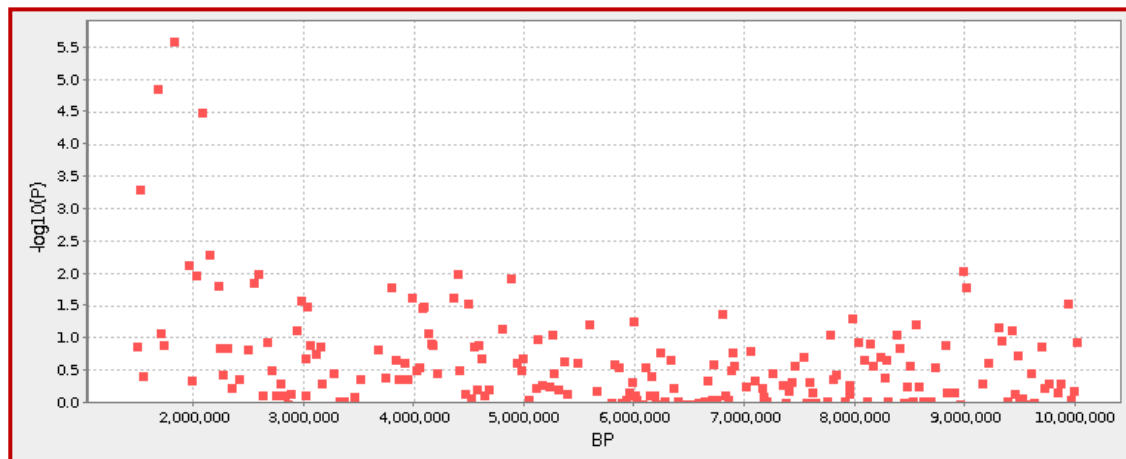


Figura 21. Cromossoma 14. BP – Pares de base; $-\log_{10}(P)$ - Corresponde ao negativo do logaritmo de base 10 do valor de p.

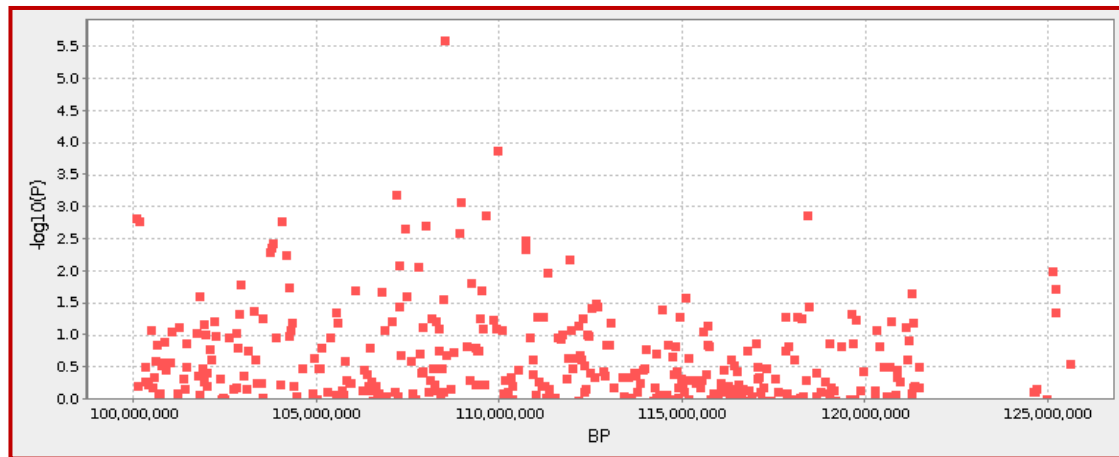


Figura 22. Cromossoma 3. BP – Pares de base; $-\log_{10}(P)$ - Corresponde ao negativo do logaritmo de base 10 do valor de p.

DISCUSSÃO E CONCLUSÃO

Como dito anteriormente na introdução deste Relatório Final de Estágio os estudos de associação genética testam a correlação entre um fenótipo e as variações genéticas para identificar genes candidatos ou regiões do genoma que contribuem especificamente para o fenótipo.

Na Figura 18, na qual temos um *Manhattan plot* com os resultados de associação dos SNPs (por cromossoma) com os valores genéticos da gordura no leite (representados pelo negativo do logaritmo do valor de p – indicam a significância do efeito desses SNPs na expressão da característica), vemos que existem dois cromossomas (cromossomas 3 e 14) que se destacam pelo valor de p que apresentam ($p < 0,00001$). Ao pesquisar na literatura identificou-se o gene DGAT1, descrito por vários autores como envolvido na quantidade de gordura no leite e também com a deposição de gordura em bovinos de carne (Vechetini, 2007), e a sua localização também está descrita como sendo no cromossoma 14 (Grisart *et al.*, 2001) o que vai ao encontro do resultado obtido.

Entretanto é necessário cautela na interpretação deste resultado. Na verdade o valor de p não pode ser interpretado como medida de magnitude da associação entre duas variáveis, mas antes apenas uma probabilidade que uma associação seja um achado falso-positivo decorrente do acaso (Botelho *et al.*, 2008). Contudo, dado o valor de p pequeno e de ser coerente com os estudos até então publicados, permite-nos ter alguma confiança na sua significância.

Também sobre a localização de genes associados a percentagem de gordura no leite temos o trabalho de Wibowo *et al.* (2008) que descreve, nos bovinos de leite, a região do cromossoma 14 compreendida entre 0-10 Mb (mega pares de bases) como uma região com alta densidade de QTLs (um total de 56 *loci*), sendo estes *loci* principalmente associados às características de produção de leite. Com o gráfico da Figura 19, que também se trata de um *Manhattan plot* onde mostra os valores de p pela localização (em pares de base) de cada SNP do cromossoma 14, vemos que há alguns valores significativos abrangidos justamente nesse intervalo (0-10 Mb). Estes resultados por si só não nos permitem garantir que estamos perante o gene DGAT1, mas não deixa de ser um forte indicador e permite ter alguma confiança nos resultados obtidos. Neste sentido, a fim de se encontrar uma localização mais aproximada, fez-se um novo gráfico, Figura 21, semelhante aos anteriores obtendo-se como resultado um intervalo ainda menor entre 0 e 2 Mb, correspondendo a porção mais inicial do cromossoma 14.

Na literatura, relativamente ao cromossoma 3, vemos que neste também se encontram alguns genes que influenciam a gordura no leite (Calvo *et al.*, 2006), compatível por isso com o resultado encontrado na Figura 18.

No gráfico da Figura 20 vemos que a principal região identificada dentro do cromossoma 3, com um valor de p inferior 0,00001, compreendeu o intervalo entre 100-125 Mb. Ao fazer um novo gráfico, Figura 22, compreendido justamente neste intervalo (100-125 Mb) vemos que o valor de p com maior significância está situado entre 107 e 110 Mb. Sendo por isso um intervalo de escolha na pesquisa de QTL para a característica em estudo.

Em conclusão, genes (ou regiões cromossómicas) associados à expressão fenotípica da gordura do leite foram identificadas nos cromossomas 3 e 14. Uma pesquisa mais extensiva nestas regiões associadas poderá contribuir para o aprofundar do conhecimento dos processos biológicos envolvidos, além de possibilitar a identificação de QTLs na determinação da expressão fenotípica. A eventual descoberta destes genes poderá ser utilizados na seleção em programas de melhoramento ao selecionar os animais portadores destes alelos.

BIBLIOGRAFIA

- Anônimo (2012) “BovineSNP50 Genotyping BeadChip” **Data Sheet: DNA Analysis** Illumina Pub. No. 370-2007-029, 1-2
- Barret JC, Fry B, Maller J, Daly MJ (2005) “Haploview: analysis and visualization of LD and haplotype maps” **Bioinformatics** 21, 263-264
- Botelho F, Silva C, Cruz F (2008) “Epidemiologia explicada – O valor de prova (p)” **Acta Urológica** 25, 55-57
- Calvo JH, Martínez-Royo A, Silveri L, Floriot S, Eggen A, Marcos-Carcavilla A, Serrano M (2006) “Isolation, mapping and identification of SNPs for four genes (ACP6, CGN, ANXA9, SLC27A3) from a bovine QTL region on BTA3” **Cytgenetic Genome** 114, 39-43
- Coutinho LL, Jorge EK, Rosário MF, Moura SAMT, Ledur MC (2007) “Genômica Animal” **Congresso Brasileiro de Zootecnia** 17, Sociedade Brasileira de Zootecnia, 429-430
- Coutinho LL, Rosário MF (2010) “Biotecnologia animal” **Estudos Avançados** 24 (nº70), 124-125
- Davis GP, Danise SK (1998) “The impact of Genetic Markers on Selection” **Journal of Animal Science** 76, 2331-2339
- Eenennaam AV (2004) “DNA-Based, Marker-Assisted Selection in Beef Cattle” **Cattle Producer’s Library** CL1043, 1-4
- Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R (2001) “Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition” **Genome Research** 12, 222-231
- Kalinowski ST (2002) “Evolutionary and statistical properties of three genetic distances” **Molecular Ecology** 11, 1263
- Klug WS, Cummings MR, Spencer CA, Palladino MA (2012) “Introduction to Genetics” **Concepts of genetics** 10, Pearson Education, 1-5
- Ledur MC, Bertani GR, Nones K (2003) “Genômica nos Programas de Melhoramento Genético Avícola” **APINCO 2003: Conferência de Ciência e Tecnologia Avícolas** Embrapa, 87-105

Lewis MC, Knight J (2012) “Introduction to genetic association studies” **Cold Spring Harb Protocols** 3, 297-306

MATLAB (2010) **The MathWorks, Inc.**, 2010

Mazoyer M, Roudart L (2009) “Domesticação e domesticabilidade” **História das agriculturas no mundo: do neolítico à crise contemporânea** 3, 125

Montaldo HH, Meza-Herrera CA (1998) “Use of molecular markers and major genes in the genetic improvement of livestock” **Electronic Journal of Biotechnology** 1 (2), 84-89

Mudadu MA, Regitano LCA, Tizioto PC (2011) “Paternityhd versão 1.0 – um software para cálculo da probabilidade de exclusão de paternidade com dados de genotipagem em painel de alta densidade de SNPs em bovinos” **Embrapa Pecuária Sudeste** 103, 6

Purcell S (2010) **PLINK (1.07) Documentation**, 35

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) “PLINK: a toolset for whole-genome association and population-based linkage analysis” **American Journal of Human Genetics** 81, 559-574

Regitano LCA, Tizioto PC, Oliveira PSN, Mudadu MA, Mourão GB (2012) “Associação Genômica com macies de carne na raça Nelore” **IX Simpósio Brasileiro de Melhoramento Animal** 1-3

SAS Institute Inc., SAS/STAT® (1990) “User’s Guide” Version 6, Fourth Edition, Volume 1, Cary, NC : **SAS Institute Inc.**, 943 pp

Tang J, Lee SH, Goddard ME, Visscher PM (2010) “GCTA: a tool for genome-wide complex trait analysis” **American Journal of Human Genetics** 88(1), 76-82

Telo da Gama L (2002) “Seleção – Valor genético real e estimado; heritabilidade” **Melhoramento Genético Animal** 1, Escolar Editora, 127

Van der Werf J (2000) “Basics of Marker Assisted Selection” **QTL Course – Identifying and incorporating genetic markers and major genes in animal breeding programs** 15, 119-121

Vechetini ME (2007) “Caracterização de polimorfismos no gene DGAT1 e no gene da Leptina em uma população de novilhas nelore” **UNESP** 2-3

Wibowo TA, Gaskins CT, Newberry RC, Thorgaard GH, Michal JJ, Jiang Z (2008) “Genome Assembly Anchored QTL Map of Bovine Chromosome 14” **International Journal of Biological Sciences** 4 (6), 406-414

Wiggans GR, VanRaden PM, Cooper TA (2011) “The genomic evaluation system in the United States: Past, present, future” **Journal of Dairy Science** 94, 3202-3211